Award Number: W81XWH-12-1-0521

TITLE: Identification of a Genomic Signature Predicting for Recurrence in Early-Stage Ovarian Cancer

PRINCIPAL INVESTIGATOR: Michael J. Birrer

CONTRACTING ORGANIZATION: Massachusetts General Hospital
Boston, MA 02114

REPORT DATE: December 2015

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2015	Final	30Sept2012 - 29Sept2015

4. TITLE AND SUBTITLE

Identification of a Genomic Signature Predicting for Recurrence in Early-Stage Ovarian Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0521

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael Birrer/ mbirrer@partners.org
Andres Poveda/ apoveda@fivo.org
Gunnar Kristensen/ post@oslo-universitetssykehus.no
Tain McNeish/ i.a.mcneish@qmul.ac.uk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Massachusetts General Hospital (The General Hospital Corp)
55 Fruit St
Boston, MA 02114-2554

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The second year of the grant required the following tasks:

1. RNAseq of 400 training specimens (Months 12-18)
2. Import raw data into public databases (Months 12-18)
3. Generate preliminary gene signature through bioinformatic and statistical analysis (Months 18-24).

In year 1) we had identified 592 early-stage high-grade ovarian cancers with 5-year follow-up, clinical annotation and accurate pathological review (228 recurrent and 364 non-recurrent), 2) established a specimen repository and clinical data inventory at MGH, 3) micro-dissected and isolated RNA from 110 tumors, and 3) optimized the preparation of cDNA libraries using NuGene WT-Ovation FFPE System V2. Given the fact that RNA sequencing is in its early stage of application, and application of this technology to FFPE tissue is still being fully developed, we have been working a work-process with different Nextgen facilities to successfully apply this technology to our FFPE samples. This included sequencing a sample test of 10 tumors and comparing the sequencing results of these early stage samples with publicly available RNAseq data for early and advanced ovarian cancers. Once the SOP were set we have been able to sequence the first 100 samples of our biorepository.

15. SUBJECT TERMS
Early Stage Ovarian Cancer, genomic predictive signature, recurrence

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	10	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. **Introduction…….1**
2. **Keywords…….1**
3. **Accomplishments……..……...1**
4. **Impact……...……7**
5. **Changes/Problems...…..……N/A**
6. **Products…….……..…..…….N/A**
7. **Participants & Other Collaborating Organizations……N/A**
8. **Special Reporting Requirements……N/A**
9. **Appendices……N/A**

Introduction

All patients with high-risk early stage ovarian cancer are treated with comprehensive surgery followed by chemotherapy over a four to six month period. Yet it is clear that many of these women are cured by surgery alone. The overtreatment results from our inability to accurately identify patients who will not likely recur with surgery alone. This ultimately exposes these patients to both short and long term toxicities from chemotherapy. An objectively measurable characteristic (i.e. biomarker) that could accurately predict for ovarian cancer recurrence would be of great clinical value much like *Oncotype DX* has done for triaging early stage breast cancer patients. This ovarian biomarker would enable health care providers to provide a more tailored approach to ovarian cancer patients. We have identified a preliminary but promising genomic signature (i.e. characteristic expression of a set of genes) that can be applied to surgically attained ovarian specimens and predicts for cancer recurrence. While we do not expect this precise signature to validate, it is proof of principle that this type of genomic tool can be identified. This project proposes to generate and validate a recurrence signature for early stage ovarian cancer. A key bottleneck precluding the validation of cancer-related signatures, in general, lies in the large number of specimens needed to ensure that the signature is clinically valuable. This proposal will utilize a larger number of early stage ovarian cancer specimens obtained from an international consortium of clinical research groups to identify a genomic signature which can accurately identify patients who will suffer tumor recurrence. The stratification of patients according to risk of recurrence will allow those patients at high risk to receive more intense therapy and those at low risk to avoid chemotherapy toxicities. This will provide patients with early stage ovarian cancer a more personalized approach in addition to reducing overall costs of treatment. The identification of a recurrence signature will occur over the three years of the grant and due to our industrial collaborations, we expect the genomic signature to rapidly transition into a commercially available tool. In addition, all specimens will undergo extensive genomic analyses to generate a publically available database of genetic changes within early stage ovarian cancer to help researchers worldwide identify biomarkers that can aid early detection and inform novel targets for therapy. This will provide a unique database which will complement existing publically available genomic data. This project will leverage unique individual banks of stored specimens and associated clinical data present in the collaborating but disparate organizations. This will allow this clinically important question to be addressed and fulfill an important unmet need.

KEYWORDS: Early Stage Ovarian Cancer, genomic predictive signature, recurrence, RNAseq

Research Accomplishments

Task 1: Using international consortium, linking multiple biorepositories and securing tissue specimens (Months 1-10)

Two accomplish this task, the following has been done in year 1: 1) Obtained IRB approval from all the Consortium collaborative Institutions to receive de-identified FFPE tissues, 2) Each site has then compiled a specimen inventory, 3) Specimens were sent to MGH, and one slide from each case was sent to GOG for review by Dr. Ramirez, 4) Clinical data for all accepted

specimens were collected and we have established a tissue biorepository with related clinical database including 592 early-stage high-grade ovarian cancers with 5-year follow-up (228 recurrent and 364 non-recurrent).

Task 2: Preparing Training Set of specimens for genomic analysis (Months 4-12)

By the end of the first project period we have optimized a protocol for the macro-dissection of our FFPE samples and started extracting nucleic acids from the first 100 samples. It is important to note that during the second year of this project we secured funding from DOD through an additional grant (W81XWH-14-1-0194) that aims to analyze DNA copy number variations in the same samples. The goal of this additional project is to integrate the DNA analysis with RNAseq in order to obtain a more robust signature. We have thus developed a protocol to extract both DNA and RNA from the same samples.

Development of a standardized protocol for FFPE ribonucleic acid extraction: A standardized protocol was developed for RNA extraction. To minimize the interference from tumor stroma derived expression profile, cresyl violet guided macro-dissection was introduced to ensure at least 80% tumor cell content within the samples subjected to nucleic acid extraction. Cresyl violet forms non-covalent, easily-reversible binding to nucleic acids and allows distinguishing tumor tissue from stroma. The staining provided by Cresyl violet is comparable to traditional dyes such as hematoxylin but, unlike hematoxylin, it does not chemically modify DNA or RNA and does not interfere with downstream profiling study (Figure 1). Dual DNA/RNA extraction from FFPE sections was then carried out by sequential use of QIAGEN miRNeasy FFPE kit (217504) and QIAGEN QIAamp® DNA FFPE Tissue Kit (56404) (Figure 2A). De-paraffinized, macrodissected FFPE tissues were briefly digested with Proteinase K in Buffer

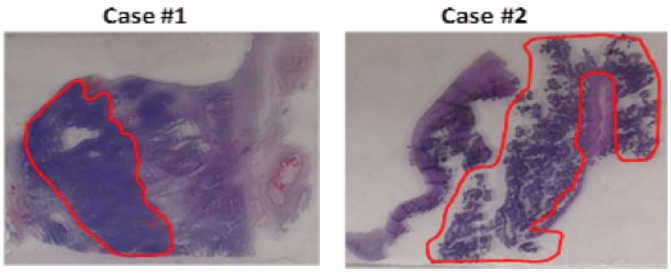

Figure 1. Cresyl violet guided macrodissection to enrich tumor component (circled by red line). In brief, deparaffinized and rehydrated FFPE tumor sections were briefly dipped into 0.5% (v/v, dissolved in 50% EtOH) cresyl violet for 30 seconds. Excessive dye was washed sequentially by 70% and 90% EtOH. Sections were then dehydrated in 100% EtOH and air-dried before macro-dissection using a sterile, RNase-free scalpel.

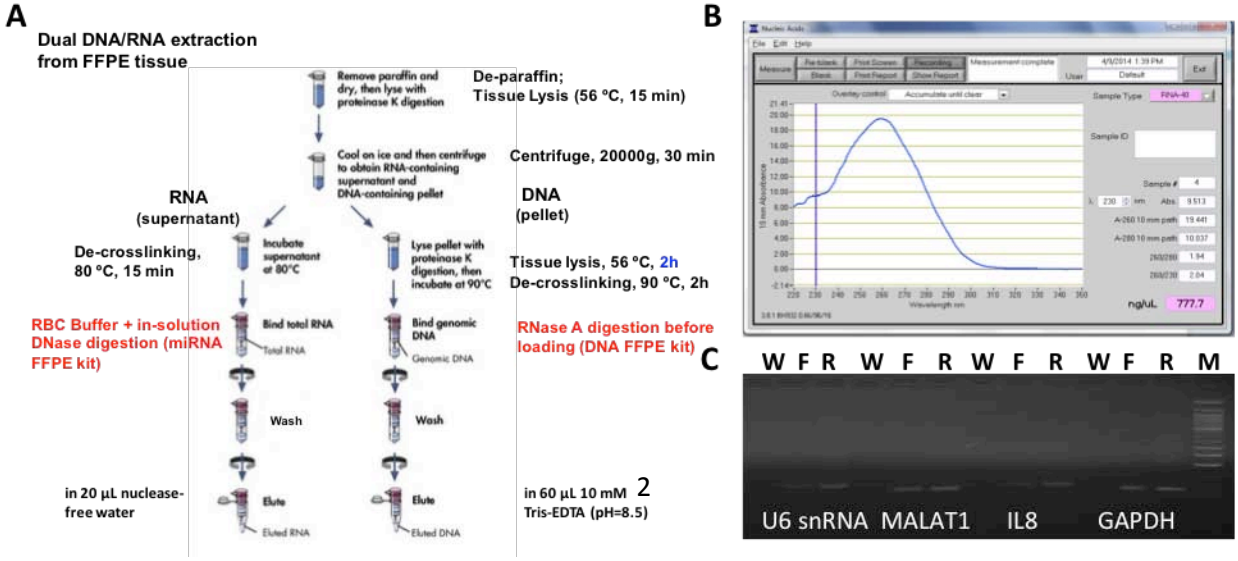

Figure 2. Flow-chart of standardized nucleic acid extraction **(A)** and related quality control **(B and C)**

PKD (QIAGEN) to release the RNA into solution while the after-digestion pellet contained primarily DNA that was processed by the QIAGEN DNA FFPE kit for a parallel project recently funded by DOD (W81XWH-14-1-0194) aimed at interrogating the genomic profiles of early-stage ovarian cancers.

After de-crosslinking to reverse the formaldehyde modification, the RNA in the solution was precipitated with increased strength of salt (guanidine HCl) and isopropanol to recover all RNA species including the mRNA and long and small non-coding RNAs (e.g. miRNA). Standard QIAGEN low-volume QIAGEN columns with preferential binding to RNA were used for RNA recovery and clean-up. The quality of the extracted RNA was checked by Nanodrop as well as qRT-PCR for various transcripts representing small RNA (U6), long non-coding RNA (MALAT1) and mRNA (IL8 and GAPDH) (Figure 2B,C).

We have finished extracting nucleic acids from all samples at the end of Year 2.

Task 3: Generation of RNAseq genomic data and generation of signature (Proposed months 12-24, accomplished in Month 32)

RNAseq from FFPE samples proved to be more challenging than what we had anticipated, mainly because we were the first group attempting to do it. Thus, instead of simply sequencing all the FFPE samples, we used 10 tumor samples (5 recurrent and 5 non recurrent) to test sequencing and establish a working protocol at three different core facilities. The sequencing results were analyzed and compared to publicly available sequences in TCGA database for early- and late-stage ovarian cancers. Based on these results we have selected the facility at the Center for Molecular Oncologic Pathology (CMOP), Dana Farber Cancer Institute (Harvard Medical School, Boston MA), and we have established library construction and RNAseq procedures that were then carried out following a Standard Operating Procedure (SOP) throughout all the samples. Thus, we have ended sequencing all samples only by the end of the third year, i.e. Month 32 and requested a no cost extension of this award to complete the studies.

Procedures for RNA extraction and sequencing:

- RNA concentration is measured by Picogreen assay (Life Tech). RNA quality control is performed on Agilent 2100 Bioanalyzer. Agilent RNA 6000 nano kit is used used for QC RNA with a minimum concentration of 5ng/uL.

- TruSeq Stranded total RNA kit is used for library preparation. The library construction protocol was optimized for degraded RNA according to guidelines from Illumina. Heating fragmentations are eliminated while only chemical fragmentations are used. Total RNA is put into RNA purification where ribosomal RNA and human mitochondrial RNA are removed by binding to magnetic micro particles with specific probes. Remaining messenger RNA and other non-coding RNA are used for generating library.

- Purified RNA is reverse transcribed to cDNA and then complementary DNA strand are synthesized to form stable double strand DNA. After 3' end adenylation, a 6 nucleotide

adaptor is ligated into the dsDNA. Libraries are enriched by 15 cycles of PCR amplification, as indicated by manufacture.

- The samples preparation is automated on BioMek FXP automation workstation (Beckman Coulter). Batches of 48 samples are processed in parallel.

This SOP has been established by the CMOP core facility and has been demonstrated to generate robust and reliable data from FFPE RNA samples. At the CMOP a similar procedure was successfully applied, in parallel with this study, to clinical prostate and lung cancer samples. For both studies paired fresh frozen (FF) and FFPE were used. The samples from both cancer types showed excellent depletion of ribosomal RNA (a major concern for non polyA selected library preparation methods), we observed less than 1% of the reads mapping to the ribosomal genes. Over 80% of the sequenced reads aligned uniquely to the human genome, a percentage comparable to the sequencing results from the frozen specimens. We observed correlations over 0.9 between the technical replicates for FFPE samples, and correlations ranging from 0.8 to 0.98 between FFPE and FF pairs. The prostate study was designed to perform biological validation of the RNA-Seq from FFPE – using paired tumor and normal specimens we were able to distinguish malignant and normal tissue using a panel of genes known to be differential expressed between these two tissue types.

Procedures for "Batch effect" control: We agreed with the core facility to prepare libraries for batches of 48 samples and sequence 4 samples at a time. Considering the relative large size of the proposed study, we noticed that the batch effect might have significant impact on data analysis for our signature development. In consultation with Dr. Victoria Wang, biostatistician from the Dana Farber Cancer Institute, we have established a two-tier of strategy to reduce the batch effect. 1) From bioinformatic prospective, standard surrogate variable analysis / principle component analysis (sva/pca) will be used to estimate artifacts introduced by factors irrelevant to biology such as sample source, sample age and technical variations. 2) We also set up SOP to minimize the technical variations during the wet-lab procedures. The latter includes: 1) performance of all extractions by only one dedicated post-doctoral fellow, and 2) tight quality control (e.g. repetitive assaying of the same sample). To perform the bioinformatics analysis of potential batch effects we have generated a fully annotated sample datasheet that records the following parameters for each sample: tumor block age, cutting-to-extraction time, tumor volume used for extraction (estimation based on number of 10µm slides), tumor purity (70 to >90% purity), DNA and RNA yield, type of stromal pattern, stromal versus tumor TILs infiltration pattern.

RNAseq analysis: Two batches of samples (96 samples total) were used to validate our sequencing protocol. These initial 96 sequences have been analyzed to: 1) better understand the efficacy and limitations of the technology 2) help reinforce the power calculation of our study and determination of the optimal ratio of recurrent versus non recurrent tumors to be used for the training stage of the study. This was important to avoid using an excessive number of samples that can be other ways used for other studies.

Receiving 80 RNAseq data from the last submission

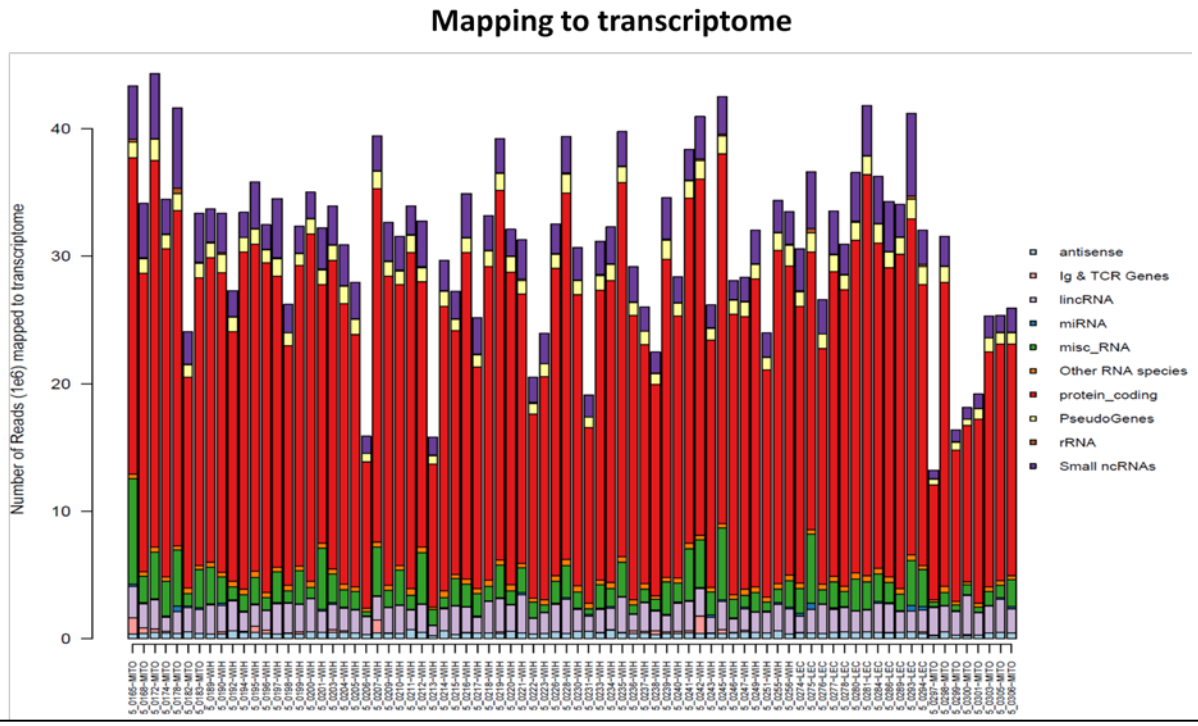

Figure 3. Graphical representation of the number of reads (Y axis) obtained for each RNA species (legend) in the first 80 FFPE tumor samples (X axis) undergoing RNAseq. While the total amount of RNAreads schanged, the proportion of mRNA (red) included over 40% of reads for each sample.

We have proven the possibility to obtain at least 40% of uniquely mapped sequences (Figure 3), thus providing the possibility to perform RNAseq on FFPE tissues. Of all samples analyzed, 8 were eliminated due to low RNA yields, and 3 fell below the threshold, determined as less than 40% unique maps, and were excluded from the analysis (Figure 4).

Figure 4. Mapping rate as a surrogate of RNAseq data QC. The majority of RNAseq data file presents a unique mapping rate of >35%. The RNAseq workflow tolerates low RNA concentration or low RNA quality (as measured by the proportion >200nt, DV>200%), but not both.

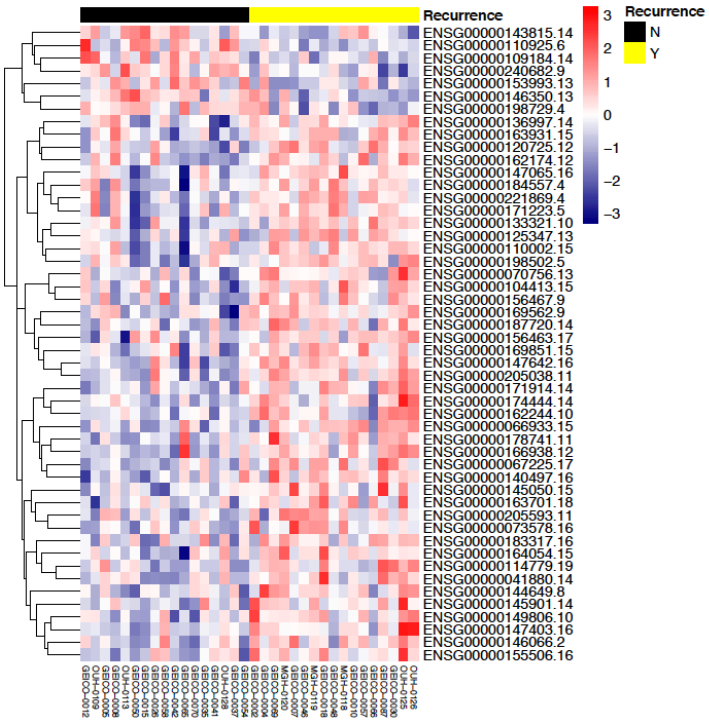

Figure 5: Clustering of the 50 top genes obtained by comparing recurrence versus no recurrence.
After obtaining normalized data for 34 serous samples, a leave-one-out cross validation was performed. In each of the 34 iterations, a different sample was reserved for validation and the rest was used for training. Limma was used to detect differentially expressed genes after transformation using Voom. When using the top 5 genes to build the classifier, it predicted recurrence status correctly 24 times out of the 34 samples (70.6%). When using the top gene as the classifier, recurrence status for 27 samples (79%) were predicted correctly.early-stage OC samples.

In addition, our analysis indicated the possibility to cluster RNAseq data from only 48 samples (Figure 5) and suggested that 384 samples (8 batches) at a ratio of 2 non recurrent tumors versus 1 recurrent tumor would have been sufficient to obtain a genomic signature distinguishing these two clinical aspects of early stage ovarian cancer.

We have then decided to sequence all samples and then divide them in a training and validation step for analysis.

Task 4: Data analysis

All samples have been sequenced and divided in training and validation sets; data analysis is ongoing It is to note, that throughout these years we have received two additional funding complementing these studies: One funding from the DOD to analyze regions of DNA amplification in the same tumors, and another funding from the Ovarian Cancer Research Fund (OCRF) to analyze expression of micro-RNA in these samples. Thus, at the end of the third year of funding we started a parallel analysis of RNAseq, DNA-CNV, and miRNAseq. To avoid any bias, we decided to analyze all all these genomics data in parallel at the same time and then integrate the results. We predict to have conclusive data by August 2017, when also the DOD award funding analysis of DNA copy number variation of the same samples, DOD-OCRP W81XWH-14-1-0194, will be completed.

Results disseminated to communities of interest: We have created a news letter that is being distributed every 2 months to communities of interest. This news letter updates the communities on the status of the project and keeps them engaged. It may be used to ask for more material. Please find attached the first version of the letter that was submitted when this project started.

IMPACT

Impact on the development of the principal discipline(s) of the project: Creation of a well annotated biorepository of early-stage tumors allows performing correlative clinical and genomic studies on these tumors that are so poorly characterized and yet significantly affect the life of so many women. Establishment of a detailed protocol for RNAseq on RNA extracted from formalin-fixed paraffin-embedded (FFPE) samples brings advancement in this novel genomic technology (RNAseq) and its broader application to cases where fresh frozen material is not available. Because most patient tumor specimens are kept as FFPE samples by the hospital, application of RNAseq to these samples allows biologic characterization of rare tumors.

Impact on other disciplines: Nothing to report

Impact on technology transfer: We anticipate that genomic discoveries in this project will have commercial application.

Impact on society beyond science and technology: Nothing to report